

Submission of Internet Archive Canada in Response to the Consultation on a Modern Copyright Framework for Artificial Intelligence and the Internet of Things

September 17, 2021

Submitted by:

Lila Bailey, Policy Counsel, and
Peter M. Routhier, Policy Fellow.

Internet Archive Canada is a not-for-profit digital library whose mission is to provide universal access to all knowledge. Over more than a decade of operations in Canada, Internet Archive Canada has digitized more than 650,000 books, government publications, and other works, a great many of which are focused on Canadian cultural heritage. Internet Archive Canada works with the Internet Archive to make these materials accessible, including in many cases for digital humanities research, machine learning, artificial intelligence, and other such applications. We write to provide our comments on a modern copyright framework for artificial intelligence.

1. Preserving a Flexible Legal Framework

In our view, a flexible legal framework is the best way to ensure copyright responds appropriately to new technological developments,¹ including artificial intelligence. Like other new technologies, artificial intelligence is a rapidly evolving, unpredictable field.² And while the consultation paper asks many of the right questions about artificial intelligence and copyright today, the answers to those questions—and the relevant questions themselves—may well change tomorrow. In the circumstances, the appropriate copyright framework for artificial intelligence must be flexible in order to keep pace with, and be in a position to respond to, unpredictable technological change.

Fortunately, Canada is already a global leader in establishing a flexible copyright framework through its user-centered approach to fair dealing.³ We were pleased to see the consultation paper recognize the role flexible copyright frameworks like fair dealing play in enabling artificial intelligence work today. And while it is true, as the paper notes, that such inquiries can be fact specific, this should not be seen as a weakness; it is their flexibility and their strength. Such flexibility is a large part of what will continue to allow Canada's copyright law—and all of its respective stakeholders—not only to keep pace with technological change, but to thrive.

¹ See von Lohmann, F. *Fair Use as Innovation Policy*. 23 Berkeley Technology Law Journal 2 (2008), available at <https://ssrn.com/abstract=1273385>.

² See, e.g., Canada's AI Research Ecosystem, available at <https://radical.vc/2021-primer-canadas-ai-research-ecosystem>.

³ See Geist, M. *The Copyright Pentology: How the Supreme Court of Canada Shook the Foundations of Canadian Copyright Law*. Ottawa: University of Ottawa Press, 2013., doi:10.1353/book.22904.

In our view, what is therefore needed is not a narrowly enumerated new exception, but rather a reaffirmation of an open, fair-use-style copyright limitation in Canada. In many ways, the recent opinion of the Supreme Court of Canada in *York University v. Access Copyright* does just that. That said, we understand the view that a lack of certainty around the governing legal framework with respect to certain artificial intelligence activities could introduce unnecessary risk into the innovation ecosystem. And as against this risk, we understand the call for a targeted AI exception. If the Government proceeds down this path, we believe that any exception should be open and flexible, incorporating and reaffirming the open, flexible nature of fair dealing in Canada today.

2. The Need for Human Review

As a library with vast digital collections, Internet Archive has already had a part to play in the development of certain machine learning and artificial intelligence techniques. For instance, Internet Archive has been a part of projects which developed machine learning techniques to identify and classify scholarly and research materials, all with a view toward providing more and better access to open scholarly materials online.⁴

One thing that we have learned from these projects is the importance of preserving an ability for human review of underlying materials.⁵ This can help guard against bad and even discriminatory results. While a detailed examination of the positives and negatives of various machine learning techniques is, of course, beyond the scope of this submission, it is enough to say that—in supervised machine learning models and beyond—it may sometimes be necessary to manually review portions of datasets both before and after ingestion.⁶ Recent research from Europe confirms that “[t]he fitness of a modern copyright system” vis-a-vis artificial intelligence must therefore be measured, in part, by whether it provides an ability to mitigate potential bad and discriminatory results by permitting “access to the original training data . . . to scrutinise [it] for mistakes, omissions or bias. . . .”⁷

Humans utilizing large datasets—whether in connection with the machine learning and artificial intelligence technologies available today, or with the as-yet-undeveloped technologies of the future—must therefore be able to review underlying materials as a part of their work. In our view, one way to do this which can provide the necessary flexibility, and which is fully compliant with the Copyright Act today, is through a technique known as controlled digital lending. Controlled digital lending is the digital

⁴ See, e.g., McNulty, J., Alvarez, S., & Langmayr, M. *Detecting Research from an Uncurated HTML Archive Using Semi-Supervised Machine Learning*. 2021 Systems and Information Engineering Design Symposium (SIEDS). <https://doi.org/10.1109/SIEDS52267.2021.9483725>, available at <https://edas.info/p28115>. See also Chang, M., Eshetu, Y., & Lemrow, C. *Supervised Machine Learning and Deep Learning Classification Techniques to Identify Scholarly and Research Content*. 2021 Systems and Information Engineering Design Symposium (SIEDS), <https://doi.org/10.1109/SIEDS52267.2021.9483792>.

⁵ There is increasing understanding that bad datasets lead to bad results; see, e.g., <https://cdt.org/ai-machine-learning/>.

⁶ See, e.g., Margoni, T. & Kretschmer, M. *A deeper look into the EU Text and Data Mining exceptions: Harmonisation, data ownership, and the future of technology* (2021) at 8-9. <https://doi.org/10.5281/zenodo.5082012>; see also McNulty, et al., *supra*.

⁷ See *id.*

equivalent of traditional library lending; it permits libraries to digitize books they own and lend them out to their patrons one at a time (just as they would with the physical book, and in place of it).⁸ It is a practice that has the support of many libraries, librarians, and others around the world, as it can make materials available in an appropriate way in view of current technologies within a recognized legal framework.⁹ Its potential application here underscores the necessity of flexible copyright frameworks for responding to the often unpredictable nature of technological change.

3. Conclusion

Internet Archive Canada greatly appreciates the government's careful consideration of these issues and the open process it has embraced with this consultation. Please do not hesitate to contact us if we can be of any further assistance.

⁸ See, e.g., <http://controlddigitallending.org>.

⁹ <https://www.ifla.org/publications/ifla-statement-on-controlled-digital-lending/>.